Syntactic Complexity of Ideal and Closed Languages [*]

Janusz Brzozowski[1] and Yuli Ye[2]

[1] David R. Cheriton School of Computer Science, University of Waterloo,
Waterloo, ON, Canada N2L 3G1
{brzozo@uwaterloo.ca }
[2] Department of Computer Science, University of Toronto,
Toronto, ON, Canada M5S 3G4,
{y3ye@cs.toronto.edu}

October 19, 2010

Abstract. The state complexity of a regular language is the number of states in the minimal deterministic automaton accepting the language. The syntactic complexity of a regular language is the cardinality of its syntactic semigroup. The syntactic complexity of a subclass of regular languages is the worst-case syntactic complexity taken as a function of the state complexity n of languages in that class. We study the syntactic complexity of the class of regular ideal languages and their complements, the closed languages. We prove that n^{n-1} is a tight upper bound on the complexity of right ideals and prefix-closed languages, and that there exist left ideals and suffix-closed languages of syntactic complexity $n^{n-1} + n - 1$, and two-sided ideals and factor-closed languages of syntactic complexity $n^{n-2} + (n-2)2^{n-2} + 1$.

Keywords: automaton, closed, complexity, ideal, language, monoid, regular, reversal, semigroup, syntactic

1 Introduction

There are two fundamental congruence relations in the theory of regular languages: the Nerode congruence [20], and the Myhill congruence [19]. In both cases, a language is regular if and only if it is a union of congruence classes of a congruence of finite index. The Nerode congruence leads to the definitions of left quotients of a language and the minimal deterministic finite automaton recognizing the language. The Myhill congruence leads to the definitions of the syntactic semigroup and the syntactic monoid of the language.

The *state complexity* of a language is defined as the number of states in the minimal deterministic automaton recognizing the language. This concept has been studied quite extensively: for surveys of this topic and lists of references

we refer the reader to [3, 31]. On the other hand, in spite of suggestions that syntactic semigroups deserve to be studied further [14, 17], relatively little has been done on the "syntactic complexity" of a regular language, which we define as the cardinality of the syntactic semigroup of the language. This semigroup is isomorphic to the semigroup of transformations of the set of states of the minimal deterministic automaton recognizing the language, where these transformations are performed by non-empty words.

The following example illustrates the significant difference between state complexity and syntactic complexity.

Example 1. The deterministic automata in Fig. 1 have the same alphabet, are all minimal, and have the same state complexity. However, the syntactic complexity of $\mathcal{A}_1$ is 3, that of $\mathcal{A}_2$ is 9, and that of $\mathcal{A}_3$ is 27.
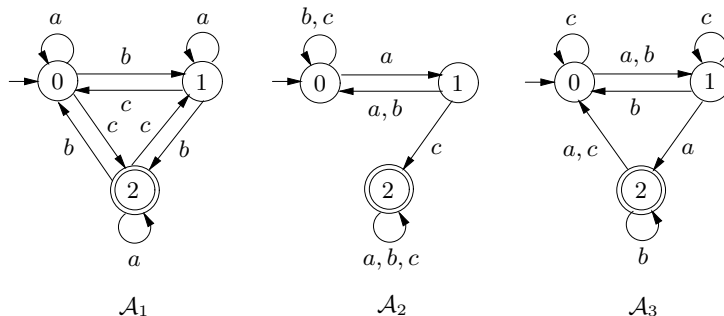
Fig. 1. Automata with various syntactic complexities.

Syntactic complexity provides an alternative measure for the complexity of a regular language. The following question then arises:

Is it possible to find upper bounds to the syntactic complexity of a regular language from its properties or from the properties of its minimal deterministic automaton?

We shed some light on this question for ideal and closed regular languages.

2   Background

This section provides a brief informal overview of the past work related to the topic of this paper. The relevant concepts will be formally defined later.

In 1938, Piccard [22] proved that two generators are sufficient to generate the set of all permutations of a set of n elements, that is, the symmetric group of degree n. The two generators can be a cyclic permutation of all the elements and

a transposition of two of the elements. References to her other early papers can be found in her books [23, 24], published in 1946, and 1957, where the problem of generators of groups is treated in detail.

In 1960, 1962, and 1963 Salomaa [26–28] studied, among other problems, the sets that can generate the set of all transformations of a set of n elements. In particular, his aim was to replace the symmetric group of degree n by smaller groups of degree n.

In 1968, Dénes [8] proved that three transformations are sufficient to generate the set of all transformations of a set S_n of n elements. One can use the two transformations that generate the symmetric group of degree n and an additional transformation that maps each of $n-1$ elements of a subset S_{n-1} of S_n to itself, and the last element to some element of S_{n-1}. Moreover, he showed that fewer than three generators are not possible. A summary of other work by Dénes on transformations can be found in [9].

In 1970, Maslov [18] dealt with the problem of generators of the semigroup of all transformations in the setting of finite automata. He pointed out that a certain ternary automaton with n states has n^n transformations. He also stated without proof that it is not possible to reach this bound with a binary automaton, and that the precise bound for the binary case is not known. He exhibited a binary automaton with n states that has at least $(n-1)^{n-1}$ transformations.

In 2002–2004, Holzer and König [12–14] studied the syntactic complexity of automata. They remarked that the syntactic complexity of a unary regular language of state complexity n is at most n, and this bound can be met. They also noted that n^n is a tight bound on the complexity of languages over alphabets Σ with $|\Sigma| \geqslant 3$. Their main contributions are in the most difficult case, that of a binary alphabet. They proved that, for $n \geqslant 3$, the function $n^n - n! + g(n)$ is an upper bound to the syntactic complexity of a binary regular language, where $g(n)$ is the Landau function. They also showed that a syntactic complexity of $n^n(1 - 2/\sqrt{n})$ can be achieved. For any prime $n \geqslant 7$, they characterized a 2-generator semigroup of maximal complexity.

In 2003, Salomaa [29] considered all the words over the alphabet Σ of a finite automaton that perform the same transformation t. In particular, he defined the length of the shortest such sequence to be the *depth with respect to Σ* of the transformation t. The *depth* of t was then defined as the maximum over all Σ that produce t. Finally, he defined the *complete depth* of a transformation to be its depth when Σ ranges over all alphabets that generate all the transformations. Many properties of the depth functions are established in this paper.

In 2003 and 2005, Krawetz, Lawrence and Shallit [17] studied the state complexity of the operation $\mathrm{root}(L) = \{w \in \Sigma^* \mid \exists n \geqslant 1$ such that $w^n \in L\}$, which is bounded from above by n^n, where n is the state complexity of L. In fact, they showed that a finite automaton with at most n^n states can be constructed to accept $\mathrm{root}(L)$, and obtained a lower bound on the state complexity of $\mathrm{root}(L)$ for binary L. For alphabets of at least three letters, they showed that the bound on the state complexity of $\mathrm{root}(L)$ can be improved to $n^n - \binom{n}{2}$.

3

3   Ideal and Closed Languages

If $w = uxv$ for some $u, v, x \in \Sigma^*$, then u is a *prefix* of w, v is a *suffix* of w, and x is a *factor* of w. A prefix or suffix of w is also a factor of w.

A language L is *prefix-convex* [1] if $u, w \in L$ with u a prefix of w implies that every word v must also be in L if u is a prefix of v and v is a prefix of w. It is *prefix-closed* if $w \in L$ implies that every prefix of w is also in L. In the same way, we define *suffix-convex* and *factor-convex*, and the corresponding closed versions.

A language $L \subseteq \Sigma^*$ is a *right ideal* (respectively, *left ideal*, *two-sided ideal*) if it is non-empty and satisfies $L = L\Sigma^*$ (respectively, $L = \Sigma^* L$, $L = \Sigma^* L \Sigma^*$). We refer to all three types as *ideal languages* or simply *ideals*.

Suffix-closed languages were studied in 1974 by Gill and Kou [11], in 1976 by Galil and Simon [10], in 1979 by Veloso and Gill [30], in 2001 by Holzer, K. Salomaa, and Yu [15], in 2009 by Kao, Rampersad, and Shallit [16] and by Ang and Brzozowski [1], and in 2010 by Brzozowski, Jirásková and Zou [6].

Left and right ideals were studied by Paz and Peleg [21] in 1965 under the names "ultimate definite" and "reverse ultimate definite events". Complexity issues of conversion of nondeterministic finite automata to deterministic finite automata in right, left, and two-sided ideals were studied in 2008 by Bordihn, Holzer, and Kutrib [2]. The closure properties of ideals were analyzed in [1]. Decision problems for various classes of convex languages, including ideals, were addressed by Brzozowski, Shallit and Xu in [7].

4   Transformations

A *transformation* of a set Q is a mapping of Q *into* itself, whereas a *permutation* of Q is a mapping of Q *onto* itself. In this paper we consider only transformations of finite sets, and we assume without loss of generality that $Q = \{0, 1, \ldots, n-1\}$. An arbitrary transformation has the form

$$t = \begin{pmatrix} 0 & 1 & \cdots & n-2 & n-1 \\ i_0 & i_1 & \cdots & i_{n-2} & i_{n-1} \end{pmatrix},$$

where $i_k \in Q$ for $0 \leqslant k \leqslant n-1$. To simplify the notation, such a transformation will often be denoted by $t : [i_0, i_1, \ldots, i_{n-2}, i_{n-1}]$, or just $[i_0, i_1, \ldots, i_{n-2}, i_{n-1}]$ if t is understood. The *identity* transformation is the mapping

$$t = \begin{pmatrix} 0 & 1 & \cdots & n-2 & n-1 \\ 0 & 1 & \cdots & n-2 & n-1 \end{pmatrix}.$$

We will consider *cycles* of length k of the following form:

$$t = \begin{pmatrix} 0 & 1 & \cdots & i-1 & \mathbf{i} & \mathbf{i+1} & \cdots & \mathbf{i+k-2} & \mathbf{i+k-1} & i+k & \cdots & n-2 & n-1 \\ 0 & 1 & \cdots & i-1 & \mathbf{i+1} & \mathbf{i+2} & \cdots & \mathbf{i+k-1} & \mathbf{i} & i+k & \cdots & n-2 & n-1 \end{pmatrix},$$

where we show in bold type the elements that are changed by t. To simplify the notation, such a cycle is represented by $(i, i+1, \ldots, i+k-1)$. A cycle of length 1 is the identity. A *singular* transformation is a transformation of the form

$$t = \begin{pmatrix} 0 & 1 & \cdots & i-1 & \mathbf{i} & i+1 & n-2 & n-1 \\ 0 & 1 & \cdots & i-1 & \mathbf{j} & i+1 & n-2 & n-1 \end{pmatrix},$$

which is denoted by $\binom{i}{j}$. The singular transformation $\binom{i}{i}$ is the identity. For $i < j$, a *transposition* is a transformation of the form

$$t = \begin{pmatrix} 0 & 1 & \cdots & i-1 & \mathbf{i} & i+1 & \cdots & j-1 & \mathbf{j} & j+1 & \cdots & n-2 & n-1 \\ 0 & 1 & \cdots & i-1 & \mathbf{j} & i+1 & \cdots & j-1 & \mathbf{i} & j+1 & \cdots & n-2 & n-1 \end{pmatrix},$$

which is denoted by (i, j), with (i, i) being the identity. A transposition is also a cycle of length 2.

A *constant* transformation is a transformation of the form

$$t = \begin{pmatrix} 0 & 1 & \cdots & n-2 & n-1 \\ i & i & \cdots & i & i \end{pmatrix},$$

and it is denoted by $\binom{Q}{i}$.

The set of all n^n transformations of a set Q is a monoid under composition of transformations, with identity as the unit element. The set of all $n!$ permutations of Q is a group, the *symmetric* group of degree n. The following facts about generators of particular semigroups are well-known:

Theorem 1 (Permutations). *The symmetric group S_n of size $n!$ can be generated by any cyclic permutation of n elements together with an arbitrary transposition. In particular, S_n can be generated by $c = (0, 1, \ldots, n-1)$ and $t = (0, 1)$.*

Theorem 2 (Transformations). *The complete transformation monoid T_n of size n^n can be generated by any cyclic permutation of n elements together with a transposition and a "returning" transformation $r = \binom{n-1}{0}$. In particular, T_n can be generated by $c = (0, 1, \ldots, n-1)$, $t = (0, 1)$ and $r = \binom{n-1}{0}$.*

5 Quotient Complexity and Syntactic Complexity

If Σ is a non-empty finite alphabet, then Σ^* is the free monoid generated by Σ, and Σ^+ is the free semigroup generated by Σ. A *word* is any element of Σ^*, and the empty word is ε. The length of a word $w \in \Sigma^*$ is $|w|$. A *language* over Σ is any subset of Σ^*. The *left quotient*, or simply *quotient,* of a language L by a word w is the language $L_w = \{x \in \Sigma^* \mid wx \in L\}$.

An equivalence relation $\sim$ on Σ^* is a *left congruence* if, for all $x, y \in \Sigma^*$,

$$x \sim y \Leftrightarrow ux \sim uy, \text{ for all } u \in \Sigma^*. \tag{1}$$

It is a *right congruence* if, for all $x, y \in \Sigma^*$,

$$x \sim y \Leftrightarrow xv \sim yv, \text{ for all } v \in \Sigma^*. \tag{2}$$

It is a *congruence* if it is both a left and a right congruence. Equivalently, $\sim$ is a congruence if

$$x \sim y \Leftrightarrow uxv \sim uyv, \text{ for all } u, v \in \Sigma^*. \tag{3}$$

For any language $L \subseteq \Sigma^*$, define the *Nerode congruence* [20] $\rightarrow_L$ of L by

$$x \rightarrow_L y \text{ if and only if } xv \in L \Leftrightarrow yv \in L, \text{ for all } u, v \in \Sigma^*. \tag{4}$$

Evidently, $L_x = L_y$ if and only if $x \rightarrow_L y$. Thus, each equivalence class of this congruence corresponds to a distinct quotient of L.

For any language $L \subseteq \Sigma^*$, define the *Myhill congruence* [19] $\leftrightarrow_L$ of L by

$$x \leftrightarrow_L y \text{ if and only if } uxv \in L \Leftrightarrow uyv \in L \text{ for all } u, v \in \Sigma^*. \tag{5}$$

This congruence is also known as the *syntactic congruence* of L. The semigroup $\Sigma^+/\leftrightarrow_L$ of equivalence classes of the relation $\leftrightarrow_L$, is the *syntactic semigroup* of L, and $\Sigma^*/\leftrightarrow_L$ is the *syntactic monoid* of L. The *syntactic complexity* $\sigma(L)$ of L is the cardinality of its syntactic semigroup. The *monoid complexity* $\mu(L)$ of L is the cardinality of its syntactic monoid. If the equivalence containing ε is a singleton in the syntactic monoid, then $\sigma(L) = \mu(L) - 1$; otherwise, $\sigma(L) = \mu(L)$.

A *(deterministic) semiautomaton* is a triple, $\mathcal{S} = (Q, \Sigma, \delta)$, where Q is a finite, non-empty set of *states*, Σ is a finite non-empty *alphabet*, and $\delta : Q \times \Sigma \rightarrow Q$ is the *transition function*. A *deterministic finite automaton* or simply *automaton* is a quintuple $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$, where Q, Σ, and δ are as defined in the semiautomaton $\mathcal{S} = (Q, \Sigma, \delta)$, $q_0 \in Q$ is the *initial state*, and $F \subseteq Q$ is the set of *final states*. A *nondeterministic finite automaton* or simply *nondeterministic automaton* is a quintuple $\mathcal{N} = (Q, \Sigma, \eta, S, F)$, where Q, Σ, and F are as defined in a deterministic automaton, $S \subseteq Q$ is the *set of initial states*, and $\eta : Q \times \Sigma \rightarrow 2^Q$ is the *transition function*.

The *ε-function* L^ε of a regular language L is $L^\varepsilon = \emptyset$ if $\varepsilon \notin L$; $L^\varepsilon = \varepsilon$ if $\varepsilon \in L$. The *quotient automaton* of a regular language L is $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$, where $Q = \{L_w \mid w \in \Sigma^*\}$, $\delta(L_w, a) = L_{wa}$, $q_0 = L_\varepsilon = L$, $F = \{L_w \mid L_w^\varepsilon = \varepsilon\}$, and $L_w^\varepsilon = (L_w)^\varepsilon$. The number of states in the quotient automaton of L is the *quotient complexity* of L. The quotient complexity is the same as the state complexity, but there are advantages to using quotients [3]. A quotient automaton can be conveniently represented by *quotient equations* [4]:

$$L_w = \bigcup_{a \in \Sigma} aL_{wa} \cup L_w^\varepsilon, \tag{6}$$

where there is one equation for each distinct quotient L_w.

In terms of automata, each equivalence class $[w]_{\rightarrow_L}$ of $\rightarrow_L$ is the set of all words w that take the automaton to the same state from the initial state. In terms of quotients, it is the set of words w that can all be followed by the same quotient L_w.

In terms of automata, each equivalence class $[w]_{\leftrightarrow_L}$ of the syntactic congruence is the set of all words that perform the same transformation on the set of states.

6

The *transformation semigroup* (respectively, *transformation monoid*) of an automaton is the set of transformations performed by words of Σ^+ (respectively, Σ^*) on the set of states. The transformation semigroup (monoid) of the quotient automaton of L is isomorphic to the syntactic semigroup (monoid) of L.

Proposition 1. *For any language L with $\kappa(L) = n$, we have $n-1 \leqslant \sigma(L) \leqslant n^n$.*

Proof. Since every state other than the initial state has to be reachable from the initial state by a non-empty word, there must be at least $n - 1$ transformations. If $\Sigma = \{a\}$ and $L = a^{n-1}a^*$, then $\kappa(L) = n$, and $\sigma(L) = n - 1$. Thus the lower bound $n-1$ is achievable. It is evident that n^n is an upper bound, and by Theorem 2 this upper bound is achievable if $|\Sigma| \geqslant 3$. □

If one of the quotients of L is $\emptyset$ (respectively, ε, Σ^*, Σ^+), then we say that L *has* $\emptyset$ (respectively, ε, Σ^*, Σ^+). A quotient L_w of a language L is *uniquely reachable (ur)* [3] if $L_x = L_w$ implies that $x = w$. If L_{wa} is uniquely reachable for $a \in \Sigma$, then so is L_w. Thus, if L has a uniquely reachable quotient, then L itself is uniquely reachable by the empty word, *i.e.*, the minimal automaton of L is *non-returning*.

Theorem 3 (Special Quotients). *Let L be any language with $\kappa(L) = n$.*
1. If L has $\emptyset$ or Σ^*, then $\sigma(L) \leqslant n^{n-1}$.
2. If L has ε or Σ^+, then $\sigma(L) \leqslant n^{n-2}$.
3. If L is uniquely reachable, then $\sigma(L) \leqslant (n - 1)^n$.
4. If L_a is uniquely reachable for some $a \in \Sigma$, then $\sigma(L) \leqslant 1 + (n - 2)^n$.
Moreover, these effects are cumulative as shown in Table 1.

Table 1. Upper bounds on syntactic complexity for languages with special quotients.

$\emptyset$	Σ^*	ε	Σ^+		L is ur	L_a is ur
$\checkmark$				n^{n-1}	$(n-1)^{n-1}$	$1 + (n-3)^{n-2}$
	$\checkmark$			n^{n-1}	$(n-1)^{n-1}$	$1 + (n-3)^{n-2}$
$\checkmark$		$\checkmark$		n^{n-2}	$(n-1)^{n-2}$	$1 + (n-4)^{n-2}$
	$\checkmark$		$\checkmark$	n^{n-2}	$(n-1)^{n-2}$	$1 + (n-4)^{n-2}$
$\checkmark$	$\checkmark$			n^{n-2}	$(n-1)^{n-2}$	$1 + (n-4)^{n-2}$
$\checkmark$	$\checkmark$		$\checkmark$	n^{n-3}	$(n-1)^{n-3}$	$1 + (n-5)^{n-2}$
$\checkmark$	$\checkmark$	$\checkmark$		n^{n-3}	$(n-1)^{n-3}$	$1 + (n-5)^{n-2}$
$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	n^{n-4}	$(n-1)^{n-4}$	$1 + (n-6)^{n-2}$

Proof. 1. Since $\emptyset_a = \emptyset$ for all $a \in \Sigma$, there are only $n - 1$ states in the quotient automaton with which one can distinguish two transformations. Hence there are at most n^{n-1} such transformations. If L has Σ^*, then $\Sigma^*_a = \Sigma^*$, for all $a \in \Sigma$, and the same argument applies.

2. Since $\varepsilon_a = \emptyset$ for all $a \in \Sigma$, L has $\emptyset$ if L has ε. Now there are two states that do not contribute to distinguishing among different transformations. Dually, $\Sigma_a^+ = \Sigma^*$ for all $a \in \Sigma$, and the same argument applies.

3. If L is uniquely reachable then $L_w = L$ implies $w = \varepsilon$. Thus L does not appear as a result of any transformation by a word in Σ^+, and there remain only $n - 1$ choices for each of the n states.

4. If L_a is uniquely reachable, then so is L. Hence L never appears as a result of a transformation by a word in Σ^+, and L_a appears only in one transformation. Therefore there can be at most $(n-2)^n$ other transformations. $\qquad\square$
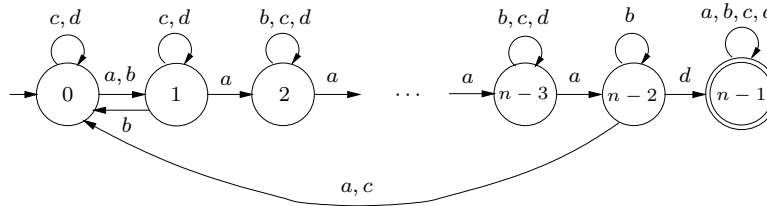
6 Right Ideals and Prefix-Closed Languages

In this section we characterize the syntactic complexity of right ideals. The automaton defined below plays an important role in this theory.

Definition 1. *For $n \geqslant 4$, define the automaton*

$$\mathcal{A}_n = (\{0, 1, \ldots, n-1\}, \{a, b, c, d\}, \delta, 0, \{n-1\}),$$

where $a = (0, 1, \ldots, n-2)$, $b = (0, 1)$, $c = \binom{n-2}{0}$, and $d = \binom{n-2}{n-1}$. The transition function δ is then defined using these transformations. The automaton so defined accepts a right ideal and is minimal; it is depicted in Fig. 2.

Fig. 2. Automaton $\mathcal{A}_n$ of a right ideal with n^{n-1} transformations.

Theorem 4 (Right Ideals and Prefix-Closed Languages)). *Let $L \subseteq \Sigma^*$ have quotient complexity n. If L is a right ideal or a prefix-closed language, then the syntactic complexity of L is less than or equal to n^{n-1}. Moreover, the bound is tight for $n = 1$ if $|\Sigma| \geqslant 1$, for $n = 2$ if $|\Sigma| \geqslant 2$, for $n = 3$ if $|\Sigma| \geqslant 3$, and for $n \geqslant 4$ if $|\Sigma| \geqslant 4$.*

Proof. Since every prefix-closed language other than Σ^* is the complement of a right ideal, and complementation preserves syntactic complexity, it suffices to consider only right ideals.

If L is a right ideal, then it has Σ^* as a quotient. By Theorem 3, we have $\sigma(L) \leqslant n^{n-1}$.

Next we prove that the language $L = L(\mathcal{A}_n)$ accepted by the automaton of Fig. 2 meets this bound. Consider any transformation t of the form

$$t = \begin{pmatrix} 0 & 1 & 2 & \cdots & n-3 & n-2 & n-1 \\ i_0 & i_1 & i_2 & \cdots & i_{n-3} & i_{n-2} & n-1 \end{pmatrix},$$

where $i_k \in \{0, 1, \ldots, n-1\}$ for $0 \leqslant k \leqslant n-2$. There are two cases:

1. Suppose $i_k \neq n-1$ for all k, $0 \leqslant k \leqslant n-2$. By Theorem 2, since all the images of the first $n-1$ states are in the set $\{0, 1, \ldots, n-2\}$, transformation t can be performed by $\mathcal{A}_n$.
2. If $i_h = n-1$ for some h, $0 \leqslant h \leqslant n-2$, then by the pigeon-hole principle, there exists some j, $0 \leqslant j \leqslant n-2$ such that $i_k \neq j$ for all k, $0 \leqslant k \leqslant n-2$.

Define i'_k for all $0 \leqslant k \leqslant n-2$ as follows:

$$i'_k = \begin{cases} j, & \text{if } i_k = n-1; \\ i_k, & \text{if } i_k \neq n-1. \end{cases}$$

Then let

$$s = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & n-3 & n-2 & n-1 \\ i'_0 & i'_1 & i'_2 & i'_3 & \cdots & i'_{n-3} & i'_{n-2} & n-1 \end{pmatrix},$$

Also, let $r = (j, n-2)$. Since all the images of the first $n-1$ states in s and r are in the set $\{0, 1, \ldots, n-2\}$, by Theorem 2, s and r can be performed by $\mathcal{A}_n$.

We show now that $t = srdr$, which implies that t can also be performed by $\mathcal{A}_n$. If t maps k to $n-1$, then s maps k to j, r maps j to $n-2$, d maps $n-2$ to $n-1$, and r maps $n-1$ to $n-1$. If t maps k to $n-2$, then s maps k to $n-2$, r maps $n-2$ to j, d maps j to j, and r maps j to $n-2$. If t maps k to $i_k < n-2$, then so does $srdr$. Hence in all cases the mapping performed by t is the same as that of $srdr$.

Since there are n^{n-1} transformations like t, $L(\mathcal{A}_n)$ meets the bound.

Now we consider the values $n \leqslant 5$. The bounds claimed below have all been verified by a computer program.

n=1: There is only one type of right ideal with $n = 1$, namely $L = \Sigma^*$, and its syntactic complexity is $\sigma(L) = 1$. Thus the bound $1^0 = 1$ is tight for $|\Sigma| \geqslant 1$.
n=2: If $|\Sigma| = 1$, there is only one right ideal, $L = aa^*$, and $\sigma(L) = 1$.
  If $|\Sigma| = 2$, then $b^*a(a+b)^*$ meets the bound $2^1 = 2$ of the theorem.
n=3: If $|\Sigma| = 1$, there is only one right ideal, $L = aaa^*$, and $\sigma(L) = 2$.
  For $n = 3$, inputs a and b of the automaton of Fig. 2 coincide.
  If $|\Sigma| = 2$, we have verified that $\sigma(L) \leqslant 7$ for all right ideals, and the language of $\mathcal{A}_3$ restricted to input alphabet $\{a, d\}$ meets the bound 7.
  If $|\Sigma| = 3$, then the language of $\mathcal{A}_3$ restricted to input alphabet $\{a, c, d\}$ meets the bound $3^2 = 9$ of the theorem.

n= 4: If $|\Sigma| = 1$, there is only one right ideal, $L = aaaa^*$, and $\sigma(L) = 3$.
For $|\Sigma| = 2$, we have verified that $\sigma(L) \leqslant 31$ for all right ideals L. The bound is reached with the inputs $a : [1, 2, 0, 3]$ and $b : [1, 0, 3, 3]$.
For $|\Sigma| = 3$, we have verified that $\sigma(L) \leqslant 61$ for all right ideals L, and $\mathcal{A}_4$ restricted to input alphabet $\{a, c, d\}$ meets this bound.

n=5: For $|\Sigma| = 2$, we have verified that $\sigma(L) \leqslant 167$ for all right ideals L. The bound is reached with the inputs $a : [0, 1, 0, 2, 4]$ and $b : [1, 3, 2, 4, 4]$, or with $a : [0, 0, 1, 2, 4]$ and $b : [2, 3, 0, 4.4]$. For $|\Sigma| = 3$, we have verified that $\sigma(L) \leqslant 545$ for all right ideals. The bound is reached with the inputs $a : [0, 0, 1, 3, 4]$, $b : [2, 0, 3, 1, 4]$, and $c : [3, 1, 2, 4, 4]$. □

Table 2 summarizes our result for right ideals. All the numbers shown are tight upper bounds. In general, there are many solutions with the same complexity.

Table 2. Syntactic complexity bounds for right ideals.

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	...	$n = n$
$\|\Sigma\| = 1$	**1**	**1**	**2**	**3**	**4**	...	**n − 1**
$\|\Sigma\| = 2$	−	**2**	**7**	**31**	**167**	...	
$\|\Sigma\| = 3$	−	−	**9**	**61**	**545**	...	
$\|\Sigma\| = 4$	−	−	−	**64**	**625**	...	**n^{n-1}**

It is interesting to note that for our right ideal L with maximal syntactic complexity, the reverse language has maximal state complexity. Recall that the reverse w^R of a word w is defined inductively by $\varepsilon^R = \varepsilon$, $(au)^R = u^R a$. The reverse of a language L is $L^R = \{w^R \mid w \in R\}$. It was shown in [5] that the reverse of a right ideal with n quotients has at most 2^{n-1} quotients, and that this bound can be met by a binary automaton. We now prove that automaton $\mathcal{A}'_n$, which is $\mathcal{A}_n$ restricted to inputs a and d, is another example of a binary automaton that meets the 2^{n-1} bound for reversal of right ideals. The nondeterministic automaton $\mathcal{N}_n$ obtained by reversing $\mathcal{A}'_n$ is shown in Fig. 3.

Theorem 5 (Reverse of Right Ideal). *The reverse of the right ideal $L(\mathcal{A}'_n)$ has 2^{n-1} quotients.*

Proof. Let Z be the set of words of the form $w = d(ae_{j-1})(ae_{j-2}) \cdots (e_1 a)(e_0)$, where $0 \leqslant j \leqslant n-2$, $e_i \in \{\varepsilon, d\}$ for $1 \leqslant i \leqslant j$. In the subset construction applied to $\mathcal{N}_n$, word da^j reaches $\{n-2-j, n-1\}$. For $1 \leqslant i < j$, the set of states reached by w includes state $n - 2 - i$ if and only if $e_i = d$. Thus each word w reaches states $n-1$, $n-2-j$, and a different subset of $\{n-2, n-3, \ldots, n-2-(j-1)\}$. There are 2^j such subsets. As j ranges from 0 to $n-2$, we get $2^0 + 2^1 + \cdots + 2^{n-2}$

10
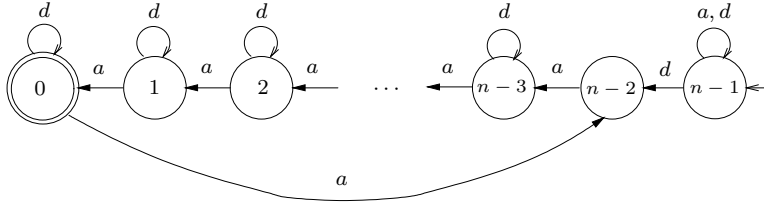
Fig. 3. Nondeterministic automaton of the reverse of a right ideal.

different subsets. Adding the subset $\{n-1\}$ reached by ε, we get 2^{n-1} reachable subsets, each containing state $n-1$.

Let $K = L^R$. The only state accepting da^{n-2} is $n-1$ reached by ε. If S and T are two different subsets of $\{0, \ldots, n-1\}$ reachable by u and v, respectively, and $i \in S \setminus T$, then $a^{n-2-i} \in K_u \setminus K_v$. Hence all the words in $\{\varepsilon\} \cup Z$ are pairwise distinguishable, and $K = L^R$ has 2^{n-1} distinct quotients. $\square$

7 Left Ideals and Suffix-Closed Languages

We provide strong support for the following conjecture about left ideals and suffix-closed languages:

Conjecture 1 (Left Ideals and Suffix-Closed Languages). *If L is a left ideal or a suffix-closed language with quotient complexity $\kappa(L) = n \geqslant 1$, then its syntactic complexity is less than or equal to $n^{n-1} + n - 1$.*

We show in this section that this complexity can be reached. Since every suffix-closed language other than Σ^* is the complement of a left ideal and complementation preserves syntactic complexity, it suffices to consider only left ideals. Before attacking the conjecture itself, we prove some auxiliary results.

First we recall a result of Restivo and Vaglica [25]. Consider a semiautomaton $\mathcal{S} = (P \cup \{0\}, \Sigma, \delta)$, where 0 is a sink state, meaning that $\delta(0, a) = 0$ for all $a \in \Sigma$, and P is strongly connected. Such a semiautomaton is *uniformly minimal* if the automaton $\mathcal{A} = (P \cup \{0\}, \Sigma, \delta, q_0, F)$ is minimal for every $q_0 \in P$ and $\emptyset \not\subseteq F \subseteq P$.

One can test whether a semiautomaton is uniformly minimal with the aid of the directed *pair graph* $G = G(\mathcal{S}) = (V, E)$. The vertices of G are all the unordered pairs (p, q) of states with $p \neq q$. There is an edge from (p, q) to (r, s) if and only if $\delta(p, a) = r$ and $\delta(q, a) = s$ for some $a \in \Sigma$. Then $\mathcal{S}$ is uniformly minimal if and only if, for any pair (p, q), there is a path to $(0, r)$ for some $r \in P$.

Definition 2. *Let $n \geqslant 3$, and let $\mathcal{S}_n$ be the semiautomaton*

$$\mathcal{S}_n = (\{0, \ldots, n-1\}, \{a, b, c, d, e\}, \delta),$$

where $a = (1, 2, \ldots, n-1)$, $b = (1, 2)$, $c = \binom{n-1}{1}$, $d = \binom{n-1}{0}$, and e is the uniform transformation $\binom{Q}{1}$. The state graph of $\mathcal{S}_n$ is shown in Fig. 4. For $n = 3$ inputs a and b coincide; hence here we use $\Sigma = \{b, c, d, e\}$.

11
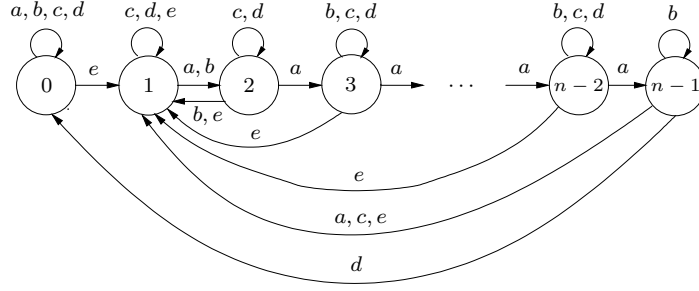
Fig. 4. Semiautomaton $\mathcal{S}_n$ with $n^{n-1} + n - 1$ transformations.

Definition 3. *Let $\Sigma' = \Sigma \setminus \{e\}$ and let $\mathcal{R}_n$ be the semiautomaton $\mathcal{R}_n = (Q, \Sigma', \delta')$, where $Q = P \cup \{0\}$, $P = \{1, \ldots, n-1\}$, and δ' is the restriction of δ to $Q \times \Sigma'$. Note that 0 is a sink state of $\mathcal{R}_n$.*

Lemma 1. *The set P is strongly connected and $\mathcal{R}_n$ is uniformly minimal.*

Proof. Since a is a cycle of the states in P, $\mathcal{R}_n$ is strongly connected.

To show that $\mathcal{R}_n$ is uniformly minimal, we construct the state-pair graph $G = G(\mathcal{R}_n)$ of $\mathcal{R}_n$, as in [25]. We need to show that for every vertex v in G, there is a path to a vertex of the form $(0, j)$, where $j \in P$.

Assume that all the unordered pairs of distinct states of $\mathcal{R}_n$ are represented as (i, j), where $i < j$. If a vertex is of the form $(0, j)$, then there is nothing to prove. If a vertex is of the form (i, j), $0 < i < j$, then applying a^{n-1-j} reaches $(i+n-1-j, n-1)$. Then d takes the pair $(i+n-1-j, n-1)$ to $(0, i+n-1-j)$. Consequently, $\mathcal{R}_n$ is uniformly minimal. $\square$

Theorem 6 (Left Ideals and Suffix-Closed Languages). *For $n \geqslant 3$, let $\mathcal{A}_n = (Q, \Sigma, \delta, 0, F)$, where $(Q, \Sigma, \delta) = \mathcal{S}_n$ of Def. 2, and F is any non-empty subset of $Q \setminus \{0\}$. Then $\mathcal{A}_n$ is minimal, and the language $L = L(\mathcal{A}_n)$ accepted by $\mathcal{A}_n$ is a left ideal and has syntactic complexity $\sigma(L) = n^{n-1} + n - 1$.*

Proof. Since semiautomaton $\mathcal{R}_n$ is uniformly minimal, automaton $\mathcal{A}_n$ is minimal for every choice of F. Hence L has n quotients.

To prove that L is a left ideal it suffices to show that, for any $w \in L$, we also have $hw \in L$ for every $h \in \Sigma$. This is obvious if $h \in \Sigma \setminus \{e\}$, since all transitions from state 0 under h lead to state 0. If $w \in L$, then w has the form $w = uev$, where $\delta(0, u) = 0$, $\delta(0, ue) = 1$, and $v \in L_e$. But $\delta(0, eue) = 1$, since $\delta(i, eue) = 1$ for all $i \in Q$, and $v \in L_e$ gives us $euev = ew \in L$. Thus L is a left ideal.

Consider any transformation t of the form

$$
t = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & n-3 & n-2 & n-1 \\ 0 & i_1 & i_2 & i_3 & \cdots & i_{n-3} & i_{n-2} & i_{n-1} \end{pmatrix},
$$

where $i_k \in \{0, 1, 2, \ldots, n-2, n-1\}$ for $1 \leqslant k \leqslant n-1$; there are n^{n-1} such transformations. We have two cases:

12

1. If $i_k \neq 0$ for all k, $1 \leqslant k \leqslant n-1$, then all the images of the last $n-1$ states are in the set $\{1, \ldots, n-1\}$. By Theorem 2, t can be performed by $\mathcal{A}_n$.
2. If $i_h = 0$ for some h, $1 \leqslant h \leqslant n-1$, then there exists some j, $1 \leqslant j \leqslant n-1$ such that $i_k \neq j$ for all k, $1 \leqslant k \leqslant n-1$.

Define i'_k for all $1 \leqslant k \leqslant n-1$ as follows:

$$
i'_k = \begin{cases} j, & \text{if } i_k = 0; \\ i_k, & \text{if } i_k \neq 0. \end{cases}
$$

Let

$$
s = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & n-3 & n-2 & n-1 \\ 0 & i'_1 & i'_2 & i'_3 & \cdots & i'_{n-3} & i'_{n-2} & i'_{n-1} \end{pmatrix},
$$

and $r = (j, n-1)$. By Theorem 2, s and r can be performed by $\mathcal{A}_n$.

Now consider $srdr$. If t maps k to 0, then s maps k to j, r maps j to $n-1$, d maps $n-1$ to 0, and r maps 0 to 0. If t maps k to $n-1$, then s maps k to $n-1$, r maps $n-1$ to j, d maps j to j, and r maps j to $n-1$. Finally, if t maps k to an element other than 0 or $n-1$, then $srdr$ maps k to the same element. Hence we have $t = srdr$, and t can be performed by $\mathcal{A}_n$ as well.

Now consider any transformation t that maps all the states to some state $j \neq 0$; there are $n-1$ such transformations. We have two cases:

1. If $j = 1$, then $t = e$; therefore t can be performed by $\mathcal{A}_n$.
2. Otherwise, let $s = (1, j)$. By Theorem 2, s can be performed by $\mathcal{A}_n$. Since $t = es$, t can also be performed by $\mathcal{A}_n$ as well.

In summary, the syntactic complexity of $L(\mathcal{A}_n)$ is $n^{n-1} + n - 1$. □

Since inputs a and b of automaton $\mathcal{A}_3$ coincide, we omit a. Table 3 shows the transition table of $\mathcal{A}_3$ and its $3^2 + 2 = 11$ transformations. We will show that 11 is indeed the maximal bound for $n = 3$, but we require more properties of left ideals.

Table 3. The eleven transformations of automaton $\mathcal{A}_3$ of a left ideal.

	b	c	d	e	bb	bd	cb	db	eb	bdb	cbd
0	0	0	0	1	0	0	0	0	2	0	0
1	2	1	1	1	1	0	2	2	2	0	0
2	1	1	0	1	2	1	2	0	2	2	0

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be the quotient automaton of a left ideal. For every word $w \in \Sigma^*$, consider the sequence $q_0 = p_0, p_1, p_2 \ldots$ of states obtained by applying powers of w to the initial state q_0, that is, let $p_i = \delta(q_0, w^i)$. Since

$\mathcal{A}$ has n states, we must eventually have a repeated state in that sequence, that is, we must have some i and $j > i$ such that $p_0, p_1, \ldots, p_i, p_{i+1}, \ldots p_{j-1}$ are distinct and $p_j = p_i$. The sequence $q_0 = p_0, p_1, \ldots, p_i, p_{i+1}, \ldots p_{j-1}$ of states with $p_j = p_i$ is called the *behavior of w on $\mathcal{A}$*, and the integer $j - i$ is the *period* of that behavior. We will use the notation $\langle p_0, p_1, \ldots, p_i, p_{i+1}, \ldots p_{j-1}; p_j = p_i \rangle$ for such behaviors. If the period of w is 1, then its behavior is *aperiodic*; otherwise, it is *periodic*.

Lemma 2. *If $\mathcal{A}$ is the quotient automaton of a left ideal L, then the behavior of every word $w \in \Sigma^*$ is aperiodic. Moreover, L does not have the empty quotient.*

Proof. Suppose that w has the behavior $\langle q_0 = p_0, p_1, \ldots, p_i, p_{i+1}, \ldots p_{j-1}; p_j = p_i \rangle$, where $j - i \geqslant 2$; then $j - 1 \geqslant i + 1$. Since $\mathcal{A}$ is minimal, states p_i and p_{j-1} must be distinguishable, say by word $x \in \Sigma^*$. If $w^i x \in L$, then $w^{j-1} x = w^i w^{j-i-1} x = w^{j-i-1}(w^i x) \notin L$, contradicting the assumption that L is a left ideal. If $w^{j-1} x \in L$, then $w^j x = w(w^{j-1} x) \notin L$, again contradicting that L is a left ideal.

For the second claim, we know that a left ideal is non-empty by definition. So suppose that $w \in L$. If L has the empty quotient, say $L_x = \emptyset$, then $xw \notin L$, which is a contradiction. $\qquad\square$

Example 2. Note that the conditions of Lemma 2 are not sufficient. For $\Sigma = \{a, b\}$, the language $L = b \cup \Sigma^* a$ satisfies the conditions, but is not a left ideal because $b \in L$ but $ab \notin L$. Its quotient automaton is shown in Fig. 5.

If the accepting state is 2 instead of 1, the language becomes $L' = \Sigma \Sigma^* b = \Sigma^* \Sigma b$, which *is* a left ideal. The languages L and L' have the same syntactic semigroup, but one is a left ideal while the other is not.
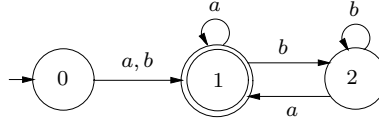
Fig. 5. Automaton of a language that is not a left ideal.

Proposition 2. *The number of transformations ruled out by Lemma 2 is*

$$\sum_{j=2}^{n} \binom{n-1}{j-1} (j-1)!(j-1)\, n^{n-j} = \sum_{j=2}^{n} \frac{(n-1)!}{(n-j)!} (j-1)\, n^{n-j}. \qquad (7)$$

Proof. Consider a behavior $\langle p_0, p_1, \ldots, p_i, p_{i+1}, \ldots p_{j-1}; p_j = p_i \rangle$ of length j. The first state, p_0, must be 0, but the set $\{p_1, \ldots p_{j-1}\}$ can be any subset of cardinality $j - 1$ of the remaining $n - 1$ states, and there are $\binom{n-1}{j-1}$ such subsets. The

states in each subset can be arranged in any order, giving $(j-1)!$ permutations. Then there are $j-1$ choices for p_j. Finally, $n-j$ states that are not part of the behavior can have n transformations each, adding the factor n^{n-j}. □

Lemma 2 provides an upper bound to the syntactic complexity of left ideals, as shown in Table 4. However, there is a large gap between this bound and the bound we can achieve, and we know that this bound cannot be reached for $n = 3$.

Table 4. Number of transformations ruled out by Lemma 2.

n	2	3	4	5	...
n^n	4	27	256	$3,125$	...
ruled out by lemma	1	10	162	$1,556$	...
an upper bound	3	17	94	$1,569$	...
$(n-1)^{n-1} + n - 1$	3	11	67	629	...

Theorem 7 (Small Left Ideals and Suffix-Closed Languages). *If* $1 \leqslant n \leqslant 3$ *and* L *is a left ideal or a suffix-closed language with* $\kappa(L) = n$, *then* $\sigma(L) \leqslant n^{n-1} + n - 1$. *Moreover, the bound is tight for* $n = 1$ *if* $|\Sigma| \geqslant 1$, *for* $n = 2$ *if* $|\Sigma| \geqslant 3$, *and for* $n = 3$ *if* $|\Sigma| \geqslant 4$.

Proof. We consider the three values of n separately. The bounds claimed below have all been verified by a computer program.

n=1: Here, there is only one type of left ideal, $L = \Sigma^*$. Thus the bound 1 holds, and is met by a^* over $\Sigma = \{a\}$.

n=2: There is only one periodic behavior $\langle p_0 = 0, p_1 = 1; p_2 = p_0 \rangle$; hence only transformation $[1,0]$ is ruled out by Lemma 2. Thus the bound 3 holds.

Now consider any left ideal L with $n = 2$. State 1 must be reachable from state 0, say by input a. By Lemma 2, we cannot have $a : [1,0]$, and so we have $a : [1,1]$.

If $\Sigma = \{a\}$, then we have the left ideal $L = a^*a$ with $\sigma(L) = 1$.
Thus $\sigma(L) = 1$ if $|\Sigma| = 1$.

If $\Sigma = \{a, b\}$, then we have three cases:
 1. If $b : [1,1]$, then $L = \Sigma^*\Sigma$ with $\sigma(L) = 1$.
 2. If $b : [0,0]$, then $L = \Sigma^*a$ with $\sigma(L) = 2$.
 3. If $b : [0,1]$, then $L = \Sigma^*a\Sigma^*$ with $\sigma(L) = 2$.
Thus $\sigma(L) \leqslant 2$ if $|\Sigma| = 2$.

If $\Sigma = \{a, b, c\}$, the language $L = \Sigma^*a(a+b)^*$ meets the bound 3.

15

n=3: For $|\Sigma| = 1$, there is only one left ideal, namely $L = \Sigma^* aa$, and it has $\sigma(L) = 2$.

For $|\Sigma| = 2$, we have verified that the number of transformations is at most 7, and the automaton with inputs $a : [001]$ and $b : [122]$ meets this bound.

For $|\Sigma| = 3$, we have verified that the number of transformations is at most 9, and the automaton $\mathcal{A}_3$ of Theorem 6 restricted to inputs $b : [0, 2, 1]$, $d : [0, 1, 0]$ and $e : [1, 1, 1]$ meets this bound.

Now consider the case $|\Sigma| = 4$. For $n = 3$, there are three types of periodic behaviors: $(p_0, p_1; p_2 = p_0)$, $(p_0, p_1, p_2; p_3 = p_0)$, and $(p_0, p_1, p_2; p_3 = p_1)$. The following ten transformations are ruled out by Lemma 2: $[1, 0, 0]$, $[1, 0, 1]$, $[1, 0, 2]$, $[1, 2, 0]$, $[1, 2, 1]$, $[2, 0, 0]$, $[2, 1, 0]$, $[2, 2, 0]$, $[2, 0, 1]$, and $[2, 2, 1]$.
There are six transformations that are not ruled out by Lemma 2 and that do not appear in Table 3, namely: $[1, 1, 0]$, $[1, 1, 2]$, $[1, 2, 2]$, $[2, 0, 2]$, $[2, 1, 1]$, and $[2, 1, 2]$. Each of these transformations, when followed by a transformation from Table 3, results in a transformation ruled out by Lemma 2:
$t_1 : [1, 1, 0]$ and $cb : [0, 2, 2]$ yield $t_1 cb : [2, 2, 0]$,
$t_2 : [1, 1, 2]$ and $db : [0, 2, 0]$ yield $t_2 db : [2, 2, 0]$,
$t_3 : [1, 2, 2]$ and $d : [0, 1, 0]$ yield $t_3 d : [1, 0, 0]$,
$t_4 : [2, 0, 2]$ and $c : [0, 1, 1]$ yield $t_4 c : [1, 0, 1]$,
$t_5 : [2, 1, 1]$ and $bdb : [0, 0, 2]$ yield $t_5 bdb : [2, 0, 0]$,
$t_6 : [2, 1, 2]$ and $bd : [0, 0, 1]$ yield $t_6 bd : [1, 0, 1]$.
All these conflicts are independent of the set of accepting states. Furthermore, each transformation not ruled out by Lemma 2 conflicts with a *different* transformation from Table 3. So at most one transformation can be chosen from each pair, showing that there cannot be more than 11 transformations for any automaton with three states. Hence the syntactic complexity of any left ideal with quotient complexity 3 is at most 11, and the example of Table 3 shows that this bound is tight. □

Table 5 summarizes our results concerning left ideals. The figures in bold type are tight upper bounds. The other complexities are achievable, but we have no proof that they are upper bounds. In general, there are many solutions with the same complexity.

The complexity 17 for $n = 4$, $|\Sigma| = 2$ is reached with the inputs $a : [1, 2, 3, 3]$ and $b : [0, 0, 1, 2]$. The complexity 25 for $n = 4$, $|\Sigma| = 3$ is met by $\mathcal{A}_4$ of Theorem 6 restricted to a, d, e. The complexity 64 for $n = 4$, $|\Sigma| = 4$ is met by $\mathcal{A}_4$ of Theorem 6 restricted to a, c, d, e.

The complexity 34 for $n = 4$, $|\Sigma| = 2$ is reached with the inputs $a : [1, 2, 3, 4, 4]$ and $b : [0, 0, 1, 2, 3]$. The complexity 65 for $n = 5$, $|\Sigma| = 3$ is met by $\mathcal{A}_5$ of Theorem 6 restricted to a, d, e. The complexity of 453 for $n = 5$, $|\Sigma| = 4$ is met by $\mathcal{A}_5$ of Theorem 6 restricted to a, c, d, e.

As was the case with right ideals, for our left ideal with maximal syntactic complexity, the reverse language has maximal state complexity, as the next result shows. This time, however, we require an alphabet of four letters.

16

Table 5. Syntactic complexities for left ideals.

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$		$n=n$
$|\Sigma|=1$	**1**	**1**	**2**	**3**	**4**	...	**n − 1**
$|\Sigma|=2$	−	**2**	**7**	17	34	...	
$|\Sigma|=3$	−	**3**	**9**	25	65	...	
$|\Sigma|=4$	−	−	**11**	64	453	...	
$|\Sigma|=5$	−	−	−	67	629	...	$n^{n-1}+n-1$

Theorem 8 (Reverse of Left Ideal). *The reverse of the left ideal accepted by automaton $\mathcal{A}_n$ of Theorem 6 restricted to $\{a, c, d, e\}$ has $2^{n-1}+1$ quotients, which is the maximum possible for a left ideal.*

Proof. Consider the subset construction applied to the nondeterministic automaton of Fig. 6. First we show that the subset $Q = \{0, 1, \ldots, n-1\}$ and all subsets of $P = Q \setminus \{0\}$ are reachable. The word $a^{n-2}e$ reaches Q, and the word $(a^{n-2}c)^{n-2}$ reaches P. Now suppose we have a set S of k elements, $S = \{i_1, i_2, \ldots, i_k\}$, where $\{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n-1\}$. To delete the jth element of S apply $a^{i_j}da^{n-1-i_j}$. Hence all subsets of P can be reached.

Note now that $a^{i-1}e$ is accepted only from state i, for $i = 1, \ldots, n-1$, and the empty word is accepted only from state 0. It follows that all subsets of P are pairwise distinguishable. □
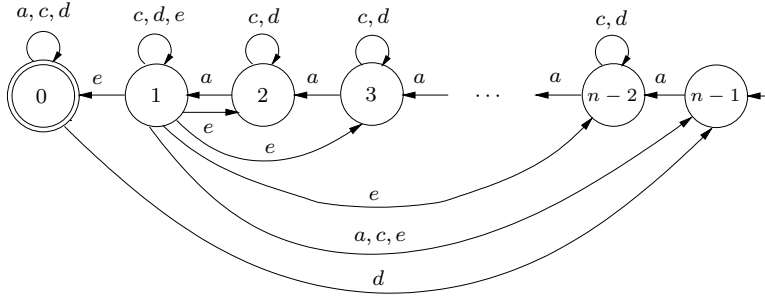
Fig. 6. Nondeterministic automaton of the reverse of a left ideal.

8 Two-Sided Ideals and Factor-Closed Languages

We now consider two-sided ideals and factor-closed languages. We provide support for the following conjecture:

Conjecture 2 (Two-Sided Ideals and Factor-Closed Languages). *If L is a two-sided ideal or a factor-closed language with quotient complexity $\kappa(L) = n \geqslant 2$, then it has syntactic complexity $\sigma(L) \leqslant n^{n-2} + (n-2)2^{n-2} + 1$.*

We show in this section that this complexity can be reached. Since every factor-closed language other than Σ^* is the complement of a two-sided ideal, and complementation preserves syntactic complexity, it suffices to consider only two-sided ideals.

For $n = 1$, the bound of the conjecture does not apply. The only two-sided ideal is $L = \Sigma^*$, and it has $\sigma(L) = 1$.

For $n = 2$ and $\Sigma = \{a, b\}$, the only two-sided ideal is $L = \Sigma^* a \Sigma^*$, and it has $\sigma(L) = 2$, which is the bound of the conjecture.

For $n = 3$ and $\Sigma = \{a, b, c\}$, the automaton with inputs $a : [1, 2, 2]$, $b : [0, 0, 2]$, and $c : [0, 1, 2]$ has $\sigma(L) = 6$, which is the bound of the conjecture.

Definition 4. *Let $n \geqslant 4$, and let $\mathcal{A}_n$ be the automaton*

$$\mathcal{A}_n = (\{0, \ldots, n-1\}, \{a, b, c, d, e, f\}, \delta, 0, \{n-1\}),$$

where $a = (1, 2, \ldots, n-2)$, $b = (1, 2)$, $c = \binom{n-2}{1}$, $d = \binom{n-2}{0}$, for $i = 0, \ldots, n-2$, $\delta(i, e) = 1$ and $\delta(n-1, e) = n-1$, and $f = \binom{1}{n-1}$. The state graph of $\mathcal{A}_n$ is shown in Fig. 7. For $n = 4$, inputs a and b coincide.
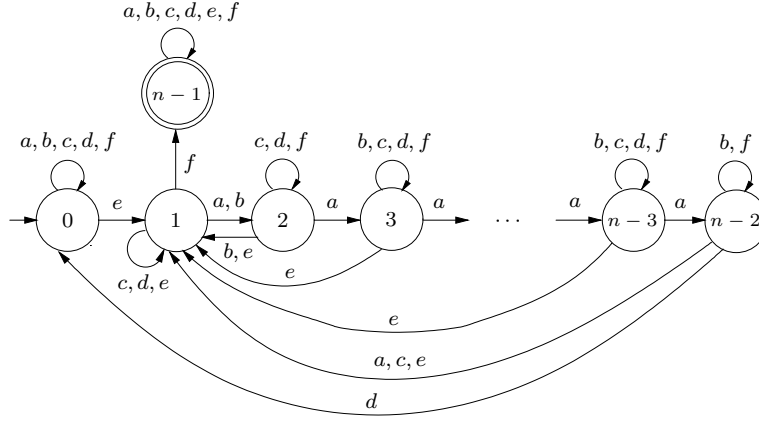
Fig. 7. Automaton $\mathcal{A}_n$ of a two-sided ideal with $n^{n-2} + (n-2)2^{n-2} + 1$ transformations.

Theorem 9 (Two-Sided Ideals and Factor-Closed Languages). *Automaton $\mathcal{A}_n$ of Fig. 7 is minimal and the language $L = L(\mathcal{A}_n)$ accepted by $\mathcal{A}_n$ is a two-sided ideal and has syntactic complexity $\sigma(L) = n^{n-2} + (n-2)2^{n-2} + 1$.*

Proof. For $i = 1, \ldots, n - 2$, state i is the only non-final state that accepts $a^{n-1-i}f$; hence all these states are distinguishable. State 0 is distinguishable from these states, because it does not accept any words in a^*f. Hence $\mathcal{A}_n$ is minimal. The proof that $\mathcal{A}_n$ is a left ideal is like that in Theorem 6. Since $L_{ef} = \Sigma^*$ is the only accepting quotient, L is a right ideal. Hence it is two-sided.

Consider any transformation t of the form

$$t = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & n-3 & n-2 & n-1 \\ 0 & i_1 & i_2 & i_3 & \cdots & i_{n-3} & i_{n-2} & n-1 \end{pmatrix},$$

where $i_k \in \{0, 1, 2, \ldots, n-2, n-1\}$ for $1 \leqslant k \leqslant n-1$; there are n^{n-2} such transformations. We have two cases:

1. If $i_k \neq n-1$ for all k, $1 \leqslant k \leqslant n-2$, then all the images of the first $n-2$ states are in the set $\{0, \ldots, n-2\}$. By Theorem 2, t can be done by $\mathcal{A}_n$.
2. If $i_h = n-1$ for some h, $1 \leqslant h \leqslant n-2$, then there exists some j, $1 \leqslant j \leqslant n-2$ such that $i_k \neq j$ for all k, $1 \leqslant k \leqslant n-2$.

Define i'_k for all $1 \leqslant k \leqslant n-2$ as follows:

$$i'_k = \begin{cases} j, & \text{if } i_k = n-1; \\ i_k, & \text{if } i_k \neq n-1. \end{cases}$$

Let

$$s = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & n-3 & n-2 & n-1 \\ 0 & i'_1 & i'_2 & i'_3 & \cdots & i'_{n-3} & i'_{n-2} & n-1 \end{pmatrix},$$

and $r = (1, j)$. By Theorem 2, s and r can be performed by $\mathcal{A}_n$.

Now consider $srfr$. If t maps k to $n-1$, then s maps k to j, r maps j to 1, f maps 1 to $n-1$, and r maps $n-1$ to $n-1$. If t maps k to 1, then s maps k to 1, r maps 1 to j, f maps j to j, and r maps j to 1. Finally, if t maps k to an element other than 1 or $n-1$, then $srfr$ maps k to the same element. Hence we have $t = srfr$, and t can be performed by $\mathcal{A}_n$ as well.

Refer to states in $\{1, \ldots, n-2\}$ as the *middle* states. Take any transformation t that maps 0 to $k \in \{1, \ldots, n-2\}$, and any middle state to either state in $\{i, n-1\}$. There are $(n-2)2^{n-2}$ such transformations. First consider any entry i that is mapped to $n-1$ by t. We can map i to $n-1$ without changing any other states. First, apply a^{n-1-i}'s to rotate all the middle states clockwise, so that i is mapped to 1, then apply f to map i to $n-1$, and then a^i to return all the states other than $n-1$ to their original positions. This is repeated for all the states that are mapped to $n-1$ by t. After this is done, apply e to replace all the middle states by 1, and apply a^{i-1} to move 1 to i. Hence t can be performed.

Finally, the constant transformation $\binom{Q}{n-1}$ is done by ef.

In summary, the syntactic complexity of the language accepted by $\mathcal{A}_n$ is at least $n^{n-2} + (n-2)2^{n-2} + 1$.

Note that 0 is mapped to a middle state 1 if and only if the input word contains an e. But every word of the form xe leaves the automaton in a state in $\{1, n-1\}$. Applying any other word can only result in a state in $\{i, n-1\}$,

for some middle state i. Hence no transformations other than the ones we have considered can be done by $\mathcal{A}_n$, and the syntactic complexity of the language accepted by $\mathcal{A}_n$ is precisely $n^{n-2} + (n-2)2^{n-2} + 1$. □

Table 6 summarizes our results for two-sided ideals. For $\Sigma = \{a, b\}$, the values are reached by the languages $\Sigma^* a^{n-1} \Sigma^*$ for $n \geqslant 2$. For $n = 4$, $|\Sigma| = 3$, the value 16 is reached by $\mathcal{A}_4$ restricted to $\{a, e, f\}$. For $|\Sigma| = 4$, the value 23 is reached by $\mathcal{A}_4$ restricted to $\{a, d, e, f\}$. For $n = 5$, $|\Sigma| = 3$, the value 47 is reached by $\mathcal{A}_5$ restricted to $\{a, e, f\}$. For $|\Sigma| = 4$, the value 90 is reached by $\mathcal{A}_5$ restricted to $\{a, d, e, f\}$. For $|\Sigma| = 5$, the value 90 is reached by $\mathcal{A}_5$ restricted to $\{a, c, d, e, f\}$.

Table 6. Syntactic complexities for two-sided ideals.

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$		$n = n$
$\|\Sigma\| = 1$	**1**	**1**	**2**	**3**	**4**	...	**n − 1**
$\|\Sigma\| = 2$	−	**2**	5	11	19	...	
$\|\Sigma\| = 3$	−	−	6	16	47	...	
$\|\Sigma\| = 4$	−	−	−	23	90	...	
$\|\Sigma\| = 5$	−	−	−	25	147	...	
$\|\Sigma\| = 6$	−	−	−	−	150	...	$n^{n-2} + (n-2)2^{n-2} + 1$

Our previous two results about reversal apply here as well.

Theorem 10 (Reverse of Two-Sided Ideal). *The reverse of the two-sided ideal accepted by automaton $\mathcal{A}_n$ of Theorem 9 restricted to $\{a, d, e, f\}$ has $2^{n-2} + 1$ quotients, which is the maximum possible for a two-sided ideal.*

Proof. Consider the subset construction applied to the nondeterministic automaton of Fig. 8. Let $P = Q \setminus \{0, n-1\}$. We will show that Q and all sets of the form $\{n-1\} \cup S$, where $S \subseteq P$ are reachable. First, Q is reached by fe. Also, $\delta(\{n-1\}, (fa)^{n-3}f) = \{n-1\} \cup P$. To remove i, $1 \leqslant i \leqslant n-2$, from any set $\{n-1\} \cup S$, apply $a^{i-1}d$; this also rotates the remaining states of P to the left by $i-1$ positions. Then apply $a^{n-2-(i-1)}$ to return the remaining states to their original positions. Hence all sets of the form $\{n-1\} \cup S$ are reachable. One verifies that all the $2^{n-2} + 1$ subsets are pairwise distinguishable. □

Despite the fact that the Myhill congruence has left-right symmetry, there are significant differences between left and right ideals. The major open problem concerning ideals is to find a better upper bound for left ideals. Also, the relation between syntactic complexity and reversal deserves further study.
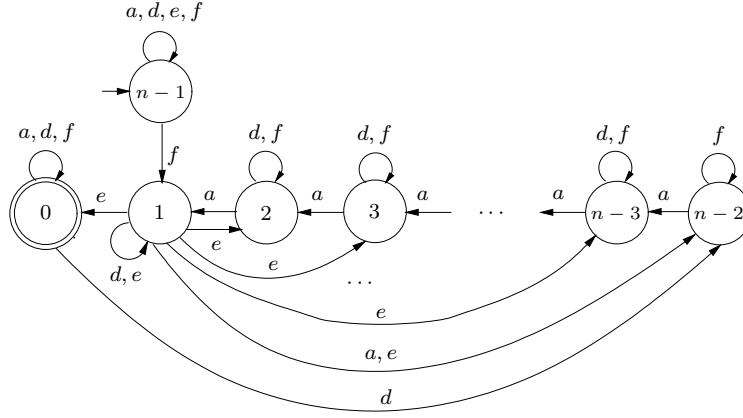
Fig. 8. Nondeterministic automaton of the reverse of a two-sided ideal.

References

1. Ang, T., Brzozowski, J.: Languages convex with respect to binary relations, and their closure properties. Acta Cybernet. **19**(2) (2009) 445–464
2. Bordihn, H., Holzer, M., Kutrib, M.: Determination of finite automata accepting subregular languages. Theoret. Comput. Sci. **410** (2009) 3209–3249
3. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. (2010) to appear.
4. Brzozowski, J.: Derivatives of regular expressions. J. ACM **11**(4) (1964) 481–494
5. Brzozowski, J., Jirásková, G., Li, B.: Quotient complexity of ideal languages. In López-Ortiz, A., ed.: Proceedings of the 9th Latin American Theoretical Informatics Symposium, (LATIN). Volume 6034 of LNCS, Springer (2010) 208–211 (Full paper at `http://arxiv.org/abs/0908.2083`).
6. Brzozowski, J., Jirásková, G., Zou, C.: Quotient complexity of closed languages. In Ablayev, F., Mayr, E.W., eds.: Proceedings of the 5th International Computer Science Symposium in Russia, (CSR). Volume 6072 of LNCS, Springer (2010) 84–95
7. Brzozowski, J., Shallit, J., Xu, Z.: Decision problems for convex languages. Information and Computation (2010) to appear.
8. Dénes, J.: On transformations, transformation semigroups and graphs. In Erdös, P., Katona, G., eds.: Theory of Graphs. Proceedings of the Colloquium on Graph Theory held at Tihany, 1966, Akadémiai Kiado (1968) 65–75
9. Dénes, J.: On generalizations of permutations: some properties of transformations. In: Actes du Colloque sur les Permutation, Paris 1972, Gauthier-Villars (1972) 117–120
10. Galil, Z., Simon, J.: A note on multiple-entry finite automata. J. Comput. Syst. Sci. **12** (1976) 350–351
11. Gill, A., Kou, L.T.: Multiple-entry finite automata. J. Comput. Syst. Sci. **9**(1) (1974) 1–19
12. Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size. In Ito, M., Toyama, M., eds.: Proceedings of the 6th International Conference on

Developments in Language Theory (DLT). Volume 2450 of LNCS, Springer (2002) 229–240

13. Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size, continued. In Ésik, Z., Fülöp, Z., eds.: Proceedings of the 7th International Conference on Developments in Language Theory (DLT). Volume 2710 of LNCS, Springer (2003) 349–360

14. Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size. Theoret. Comput. Sci. **327** (2004) 319–347

15. Holzer, M., Salomaa, K., Yu, S.: On the state complexity of k-entry deterministic finite automata. J. Autom. Lang. Comb. **6** (2001) 453–466

16. Kao, J.Y., Rampersad, N., Shallit, J.: On NFAs where all states are final, initial, or both. Theoret. Comput. Sci. **410**(47-49) (2009) 5010–5021 Available in 2003 at arXiv:math/0306416v2.

17. Krawetz, B., Lawrence, J., Shallit, J.: State complexity and the monoid of transformations of a finite set. In Domaratzki, M., Okhotin, A., Salomaa, K., Yu, S., eds.: Proceedings of the Implementation and Application of Automata, (CIAA). Volume 3317 of LNCS, Springer (2005) 213–224

18. Maslov, A.N.: Estimates of the number of states of finite automata. Dokl. Akad. Nauk SSSR **194** (1970) 1266–1268 (Russian) English translation: Soviet Math. Dokl. **11** (1970), 1373–1375.

19. Myhill, J.: Finite automata and representation of events. Wright Air Development Center Technical Report **57–624** (1957)

20. Nerode, A.: Linear automaton transformations. Proc. Amer. Math. Soc. **9** (1958) 541–544

21. Paz, A., Peleg, B.: Ultimate-definite and symmetric-definite events and automata. J. ACM **12**(3) (1965) 399–410

22. Piccard, S.: Sur les bases du group symétrique et du groupe alternant. Commentarii Mathematici Helvetici **11**(1) (1938) 1–8

23. Piccard, S.: Sur les bases du groupe symétrique et les couples de substitutions qui engendrent un groupe régulier. Librairie Vuibert, Paris (1946)

24. Piccard, S.: Sur les bases des groupes d'ordre fini. Mémoires de l'université de Neuchatel, vol. 25. Université de Neuchatel, Switzerland (1957)

25. Restivo, A., Vaglica, R.: Automata with extremal minimality conditions. In Gao, Y., Seki, S., Yu, S., eds.: Proceedings of the 14th International Conference on Developments in Language Theory (DLT). Volume 6224 of LNCS, Springer (2010) 399–410

26. Salomaa, A.: A theorem concerning the composition of functions of several variables ranging over a finite set. J. Symbolic Logic **25** (1960) 203–208

27. Salomaa, A.: Some completeness criteria for sets of functions over a finite domain. Ann. Univ. Turkuensis, Ser. AI **53** (1962)

28. Salomaa, A.: On basic groups for the set of functions over a finite domain. Ann. Acad. Scient. Fenn., Ser. A **338** (1963)

29. Salomaa, A.: Composition sequences for functions over a finite domain. Theoret. Comput. Sci. **292** (2003) 263–281

30. Veloso, P.A.S., Gill, A.: Some remarks on multiple-entry finite automata. J. Comput. Syst. Sci. **18** (1979) 304–306

31. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. **6** (2001) 221–234